УДК 004.89:004.93

Н.С. Клименко, И.Г. Герасимов
Институт проблем искусственного интеллекта МОН Украины и НАН Украины, г. Донецк
Украина, 83048, г. Донецк, ул. Артема, 118 б

Исследование эффективности бустинга в задаче текстонезависимой идентификации диктора

M.S. Klymenko, I.G. Gerasimov
Institute of Artificial Intelligence MES of Ukraine and MAS of Ukraine, c. Donetsk
Ukraine, 83048, c. Donetsk, Artema st., 118 b

Research of Boosting Efficiency for Text-Independent Speaker Identification System

М.С. Клименко, І.Г. Герасимов
Інститут проблем штучного інтелекту МОН України і НАН України, м. Донецьк
Україна, 83048, м. Донецьк, вул. Артема 118 б

Дослідження ефективності бустінгу щодо задачі текстонезалежної ідентифікації диктора

В статье рассмотрены особенности текстонезависимого подхода и основные этапы процедуры идентификации диктора. Предложена модификация типовой схемы идентификации при помощи предварительной сегментации речевого сигнала и структурирования базы моделей дикторов, а также объединения набора классификаторов акустических признаков в одно решающее правило обучением по алгоритму бустинга. Предложенные модификации повышают вероятность текстонезависимой идентификации и сокращают количество требуемых вычислений.
Ключевые слова: идентификация личности по голосу, бустинг, широкие фонетические классы, модель диктора, модели гауссовых смесей, машины опорных векторов.

In the article, features of text- independent approach and the main steps of the speaker identification procedure are considered. A modification of the standard identification scheme using pre-segmentation of the speech signal and the structuring of speaker's models base, as well as merge a set of acoustic feature classifiers into a single decision rule by boosting algorithm. The proposed modification increases the likelihood of text independent identification and reduce the number of required computations.
Keywords: speaker identification, wide phonetic classes, speaker model, Gaussian mixture models, support vector machines.

У статті розглянуто особливості текстонезалежного підходу й основні етапи процедури ідентифікації диктора. Запропоновано модифікацію типової схеми ідентифікації за допомогою попередньої сегментації мовного сигналу і структурування бази моделей дикторів, а також об'єднання набору класифікаторів акустичних ознак в одне вирішальне правило навчанням за алгоритмом бустінгу. Запропоновані модифікації підвищують ймовірність текстонезалежної ідентифікації та скорочують кількість обчислень.
Ключові слова: ідентифікація особи по голосу, бустінг, широкі фонетичні класи, модель диктора, моделі гаусових сумішей, машини опорних векторів.

Введение

Идентификация диктора представляет собой процедуру, в ходе которой необходимо производить классификацию звукового фрагмента по эталонным образам — акустическим моделям диктора, и на основании этого выносить решение о соответствии

заявленного голоса тому или иному диктору. Текстонезависимой является такая идентификация, которая выполняется по произвольному фрагменту звукозаписи речи без информации об ее текстовом содержимом.

Идентификация личности по голосу на сегодняшний день востребована широким спектром многопользовательских аппаратно-программных решений от персональной техники до систем безопасности. Определение пользователя при помощи устной речи не требует особых биометрических детекторов и непосредственного физического контакта с ними, что открывает возможность удаленной идентификации посредством сети Интернет или мобильной связи.

В текстонезависимых системах может использоваться любой фрагмент речи диктора, что делает их удобными с точки зрения пользователя, тогда как текстозависимые системы иде. Такие системы незаменимы при решении задач правоохранительной сферы: скрытая идентификация, криминалистическая идентификация, фоноучеты. Тем не менее, эта возможность усложняет реализацию текстонезависимых систем, понижает их надежность и скорость распознавания.

Метод классификации, примененный к определенным признакам в процедуре идентификации, называется классификатором. Каждый классификатор обладает некоторыми преимуществами и недостатками, по-разному реагирует на отличия в условиях обучения и распознавания, особенности голоса разных дикторов и т.д. Целесообразно использовать решения разных классификаторов, чтобы уменьшить ошибки распознавания. Учесть качество нескольких классификаторов позволяет бустинг – метод, основанный на комбинировании нескольких «слабых» классификаторов в один «сильный». Современные исследования показывают, что бустинг может успешно применяется к классификаторам графических образов [1], [2]. Данная работа заключается в адаптации алгоритма бустинга к задаче текстонезависимой идентификации диктора и оценке эффективности данного подхода.

Особенности текстонезависимого подхода и разработанной системы идентификации диктора. Текстонезависимый подход отличается использованием классификаторов, либо использующих всю совокупность признаков вместе, либо «привязывающихся» к определенным акустическим событиям (а не к последовательности признаков), таким образом самостоятельно компенсируя недостаток априорной информации о текстовом содержании аудиофрагмента. Разнообразие данных методов определяется возможностями компенсации межсессионной вариативности (т.е. отличий между записями в разное время и на разном оборудовании), канальных искажений звукозаписывающего и передающего оборудования, шума. Применение нейросетей и их различных модификаций возможно для обеих задач, представляя собой альтернативу бустингу в плане универсальности [3].

Типовая процедура текстонезависимой идентификации диктора начинается с поиска в заданном звуком отрезке фрагментов, содержащих акустические события. Найденные фрагменты разбиваются на окна, частично перекрывающиеся, и по каждому окну вычисляется вектор определенных акустических признаков. Далее в режиме обучения на основании полученных векторов признаков формируется образ – акустическая модель диктора и сохраняется в соответствующую базу данных. В режиме распознавания совокупность векторов признаков сопоставляется с имеющимися моделями дикторов и методом классификации выносится решение: идентификатор распознанного диктора, список ближайших дикторов-кандидатов или отказ распознавания в случае недостаточного количества векторов признаков или низкой меры близости ко всем моделям дикторов [4].

Разработанная система идентификации диктора имеет несколько принципиальных отличий от типовой схемы:

– автоматическая сегментация на широкие фонетические классы (ШФК);
– иерархическое структурирование базы моделей дикторов;
– одновременное использование нескольких акустических признаков.

Поскольку база данных системы содержит различные модели дикторов, основанные на нескольких акустических характеристиках, а также 2 метода классификации, структура базы моделей дикторов отличается для каждой комбинации признак-метод. Для проведения численных исследований этап принятия решения реализован в 3 модификациях: на основе 2 методов классификации и композиции «слабых» классификаторов, полученной обучением методом бустинга.

Как было отмечено, текстонезависимым методам для улучшения передачи характеристик диктора необходимо привязываться к определенным акустическим событиям. Модификацией метода гауссовых смесей и главной особенностью предложенной схемы идентификации является учет широких фонетических классов – множества, характеризующего акустическое пространство голоса диктора. Именно ШФК в данной работе представляют те акустические события, на основании которых идет дальнейшая работа методов классификации. ШФК описывают определенные фонетические события (гласные, фрикативные и т.д.), по которым в отдельности формируются акустические модели, в совокупности представляющие «комплексную» модель диктора. Существуют различные классификации звуков согласно методам их извлечения и схожести восприятия органами слуха. В этой работе по результатам численных исследований был использован состав из четырех ШФК: глухие, звонкие, сонорные и гласные. Данный состав признан оптимальным для звуков речи русского языка по параметрам межклассового и внутриклассового разбросов векторов признаков в ходе проведенного численного исследования [5].

Еще одной особенностью предложенной схемы идентификации, использующей модифицированный нами метод на основе гауссовых смесей, является структурирование моделей дикторов, занесенных в базу данных системы. Модели сохраняются с иерархическими связями между собой и фоновыми моделями групп дикторов, расположенных глубже (ниже по иерархии). Фоновая модель является аналогией модели диктора со следующими отличиями:

– построена методом машин опорных векторов по схеме «один против всех» (решающее правило разделяет одну модель от всех остальных);

– отдельными классами выступают низлежащие и сгруппированные по иерархии модели дикторов и их групповые модели.

Роль групповых фоновых моделей заключается в возможности прохождения по дереву от корня к листьям. Возможным становится и проведение открытой идентификации (когда дополнительно определяется наличие модели диктора в базе): по негативным результатам сравнения со всеми групповыми фоновыми моделями на текущем уровне иерархии можно утверждать об отсутствии распознаваемого диктора в базе моделей. Задача верификации не затрагивает данную иерархию, обращаясь к модели диктора по его идентификатору напрямую.

Таким образом, структурирование позволяет избежать полного перебора базы моделей дикторов, существенно ускоряя процесс закрытой идентификации, позволяя выполнять открытую идентификацию и верификацию по этой же базе моделей дикторов. Предлагаемая структура идентификации диктора призвана снизить влияние канальных искажений и увеличить качество идентификации за счет применения различных дикторонезависимых признаков с целью разбиения произвольной речи на участки, принадлежащие различным ШФК, и создания для каждого диктора множества моделей для каждого фонетического элемента.

В качестве акустических признаков в системах распознавания диктора могут использоваться различные параметры, учитывающие процессы как речеобразования (характеристики распределения частоты основного тона, коэффициенты линейного предсказания, спектр Фурье), так и восприятия речи (вейвлет-спектр, мел-частотные кепстральные коэффициенты), и их динамические характеристики. Все извлекаемые из аудиосигнала показатели не лишены недостатков: робастные параметры обладают слабыми идентификационными качествами и, наоборот, параметры, характеризующие диктора с высокой точностью, достаточно сильно чувствительны к различным факторам (темп, громкость произнесения, физическое и эмоциональное состояние человека, вид и уровень помех в акустическом и электронном канале связи, искажение речевого сигнала приемниками звука и реверберацией помещения) [4].

В данной работе акустическими признаками диктора выступают Мел-частотные кепстральные коэффициенты (Mel frequency cepstral coefficients, MFCC), зарекомендовавшие себя компактными и информативными признаками во множестве современных исследований как по распознаванию речи, так и диктора [6], [7]. Мел – психофизическая единица высоты звука, применяется главным образом в музыкальной акустике. Высота звука связана главным образом с частотой колебаний, но зависит также от уровня громкости звука и его тембра. Шкала Мел-частот аппроксимирует данные искажения субъективного восприятия высоты звука, позволяя приблизить представление звуковой информации к человеческим величинам. Наиболее информативной частью спектра становится именно полоса речевой активности 0,5 – 4 кГц.

MFCC вычисляются аналогично кепстральным коэффициентам с той разницей, что на после дискретного преобразования Фурье выполняется отображение спектра по шкале Мел-частот. В результате первые 13 коэффициентов охватывают спектральную область до 8 кГц, чего достаточно для распознавания диктора. Для учета динамической составляющей вектор признаков дополняется производными от аналогичных коэффициентов, полученных по соседним окнам, таким образом общая длина вектора признаков равна 39.

В качестве альтернативных акустических признаков использованы частота основного тона вокализованного фрагмента и пики 2-4 формант. По сравнению с MFCC эти признаки имеют меньшую вероятность верного распознавания при использовании одного и того же классификатора (исследование проводилось с помощью гауссовых смесей). Частота основного тона определяется при помощи автокорреляционной функции с постфильтрацией значений. Пики формант находятся по сглаженному спектру на основе коэффициентов линейного предсказания [5].

Первоначально в рамках данной работы для классификации использовались только гауссовы смеси. Этот классификатор представляет собой аналогию скрытой марковской модели с одним состоянием. Затем для построения групповых моделей дикторов был использован метод машин опорных векторов по схеме «один против всех». Отличие данных классификаторов принципиально: гауссовы смеси предназначены максимально точно описать одну конкретную выборку признаков каждой модели и, следовательно, решение о классификации будет принято на основании максимального соответствия тестовой выборки одной из моделей. А метод машин опорных векторов позволяет определить в многомерном пространстве признаков расположение гиперплоскости, являющейся равноудаленной от крайних (опорных) векторов двух противоположных классов, формируя при линейной неразделимости признаков искривленное пространство, добиваясь их разделимости. В силу принципиальных различий было принято решение использовать оба метода для построения множества «слабых» классификаторов в алгоритме бустинга.

Модели гауссовых смесей представляют классы в пространстве признаков в виде многомерного вероятностного распределения. Класс описывается как взвешенная сумма M нормальных распределений:

$$p(\bar{x} \mid \lambda) = \sum_{i=0}^{M} w_i p_i(\bar{x}), \quad \sum_{i=1}^{M} p_i = 1, \tag{1}$$

где x – N-мерный вектор признаков;
w_i – веса компонентов модели;
p_i – многомерные функции плотности распределения составляющих.

Таким образом, модель описывается векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели [8].

Способом оценки параметров модели является метод максимизации правдоподобия:

$$p(\bar{x} \in \lambda) = \sum_{i=0}^{M} w_i p_i(\bar{x}), \tag{2}$$

где λ – модель, относительно которой вычисляется вероятность соответствия.

Метод машин опорных векторов определяет принадлежность объектов к классам с помощью границ областей.

При классификации только по двум категориям (подход может быть расширен на любое конечное количество категорий), каждый объект классификации является вектором в N-мерном пространстве.

Метод машин опорных векторов базируется на таком постулате: наилучшая разделяющая прямая – это та, которая максимально далеко отстоит от ближайших до нее точек обоих классов. То есть задача метода машин опорных векторов состоит в том, чтобы найти такие вектор w и число w_0, чтобы ширина разделяющей полосы была максимальна. Поскольку чем шире полоса, тем увереннее можно классифицировать документы, соответственно, в методе машин опорных векторов считается, что самая широкая полоса является наилучшей. Границами полосы являются две параллельные гиперплоскости с направляющим вектором w. Точки, ближайшие к разделяющей гиперплоскости, расположены точно на границах полосы, при этом сама разделяющая гиперплоскость проходит ровно посередине полосы. Опорным вектором называется объект обучающей выборки x_i , находящийся на границе разделяющей полосы.

Для улучшения метода применяется идея расширенного пространства, которая заключается в переходе от исходного пространства признаковых описаний объектов X к новому пространству H с помощью некоторого преобразования ψ: $X \rightarrow H$. Если пространство H имеет достаточно высокую размерность, то вполне вероятно, что в нём выборка окажется линейно разделимой. При выборе признаковыми описаниями объектов векторов $\psi(x_i)$, а не векторов x_i , построение машин опорных векторов проводится точно так же. Отличие состоит в том, что скалярное произведение $\langle x, x' \rangle$ в пространстве X заменяется скалярным произведением $\langle \psi(x), \psi(x') \rangle$ в пространстве H. Спрямляющее пространство должно быть наделено скалярным произведением, следовательно, в общем случае применимо любое гильбертово пространство [9]. В данной работе в качестве ядра использована радиальная базисная функция:

$$H(x, x') = \exp(-c\|x - x'\|^2), \tag{3}$$

где c – настраиваемый параметр.

В блоке принятия решений реализован наиболее известный алгоритм бустинга AdaBoost. Он строит сильный классификатор машинного обучения по набору слабых алгоритмов машинного обучения путем многократного прохождения по обучающей выборке *(x1, y1),...,(xn, yn)* и увеличения веса примеров, на которых слабые алгоритмы H_m дают большую ошибку обучения e_m. Начальная функция распределения весовых коэффициентов является равномерной. Каждая итерация начинается с того, что находится лучший на текущем распределении D_m , слабый классификатор:

$$h_m(x) = argmin(e_m),$$
$$e_m = \sum_{i=1}^{N} D_m(i) \cdot h_m(x_i), \quad h_m(x_i) \neq y_i; \tag{4}$$

Затем выполняется нахождение коэффициента a_m и соответствующей функции f_m голосования *m*-го классификатора:

$$a_m = \frac{1}{2} \cdot \log\left(\frac{1-e_m}{e_m}\right)$$
$$f_m(x_i) = a_m h_m(x_i) \tag{5}$$

Итерация завершается пересчетом распределения весов:

$$D_{m+1}(i) = \left(\frac{D_m(i)\exp(-y_i f_m(x_i))}{Z_i}\right)$$
$$\sum_{i=1}^{N} D_{m+1}(i) = 1 \tag{7}$$

где Z_i – нормализующий коэффициент.

Таким образом, сильный классификатор представляется в виде:

$$F(x) = sign\left[\sum_{i=1}^{M} f_m(x)\right] \tag{8}$$

Численные исследования эффективности применения бустинга. На основе имеющихся разработок было построено множество «слабых» классификаторов. В него включены различные комбинации признак-классификатор на основе одного и всех широких фонетических классов. Проведены циклы бустинга множества классификаторов на обучающих выборках с разным количеством распознаваемых классов (дикторов) от 2 до 50. Звукозаписи дикторов были взяты из общедоступной базы VoxForge [11]. Из более чем 500 дикторов была получена случайная простая вероятностная выборка звукозаписей 50 дикторов. Звукозаписи представлены в виде wav-файлов с глубиной квантования 16 бит и частотой дискретизации 16 кГц. Все дикторы записаны динамическим микрофоном в спокойном эмоциональном состоянии, 42 диктора –мужчины, 8 – женщины, возраст дикторов не занесен в базу, на слух все дикторы средних лет без присутствия детских или престарелых голосов.

Обучение проводилось на фрагментах речи дикторов продолжительностью от 8 до 50 секунд. Во фрагментах каждого диктора было зафиксировано наличие акустических признаков, относящихся ко всем ШФК.

Результаты проведения бустинга следует разделить на прогрессивную оценку при изменении количества моделей дикторов для обучения и на относительную оценку влияния классификаторов. Также приведен сравнительный анализ эффективности идентификации диктора «слабыми» классификаторами и композицией классификаторов. Распознавание проводилось по фрагментам дикторов той же сессии звукозаписи продолжительностью 5 – 7 секунд.

Прогрессивная оценка позволяет отобразить влияние размера обучающей выборки на распределение весовых коэффициентов 5. ($p < 0{,}05$). На рис. 1 представлены графики изменения коэффициентов при $2 \leq N \leq 50$ для различных ШФК, используемых в данной работе.

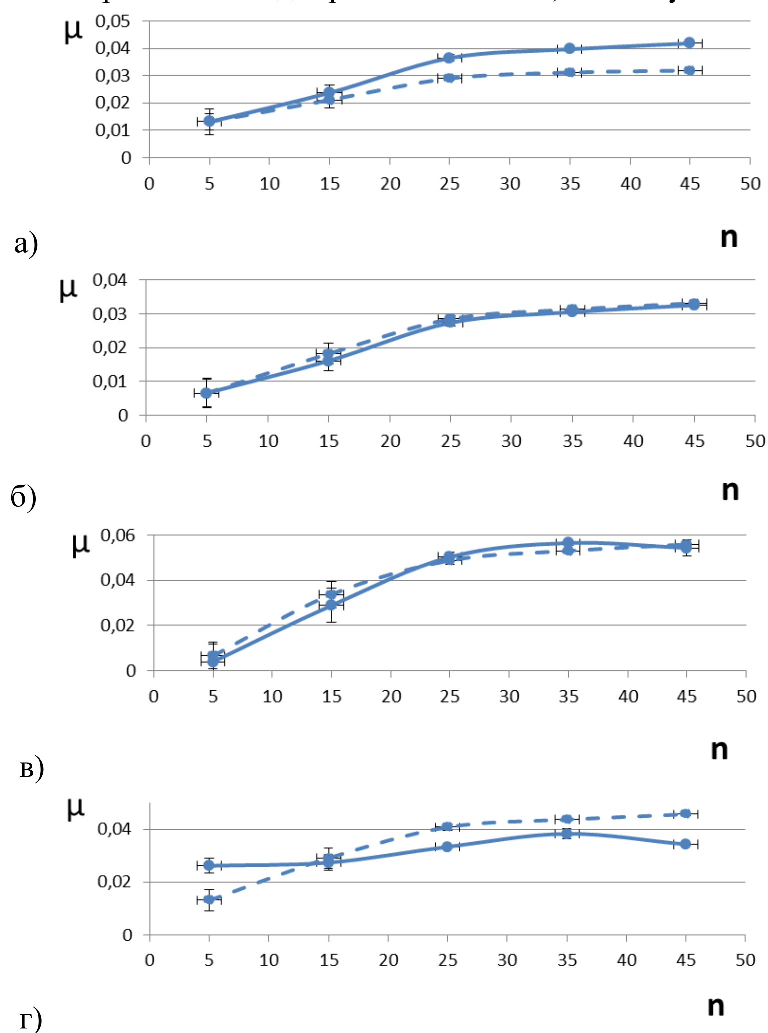


Рисунок 1 – Зависимость весовых коэффициентов (μ) от количества дикторов (n) «слабых» классификаторов с использованием различных ШФК (а – глухие, б – звонкие, в – сонорные, г – гласные), применяя мел-частотные кепстралные коэффициенты в качестве акустических признаков, а также гауссовы смеси (пунктирная линия) и машины опорных векторов (сплошная линия) в качестве методов классификации

Аналогичная зависимость для классификаторов с использованием в качестве акустических признаков частот основного тона и формантных пиков представлена на рис. 2.

Зависимость весовых коэффициентов от размера обучающей выборки (количества моделей диктора) $2 \leq N \leq 50$ позволяет установить:

– значительное уменьшение влияния классификаторов, основанных на ЧОТ, при увеличении количества моделей;

– стабильно высокое влияние классификаторов методом опорных векторов с комплексным признаками ШФК;

– из классификаторов с признаками по одному из ШФК наибольший вклад показывают вокализованные при малом количестве дикторов и сонорные с шипящими – при большем количестве дикторов.

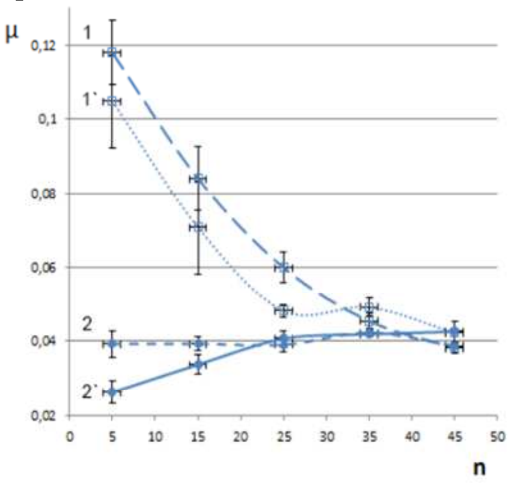
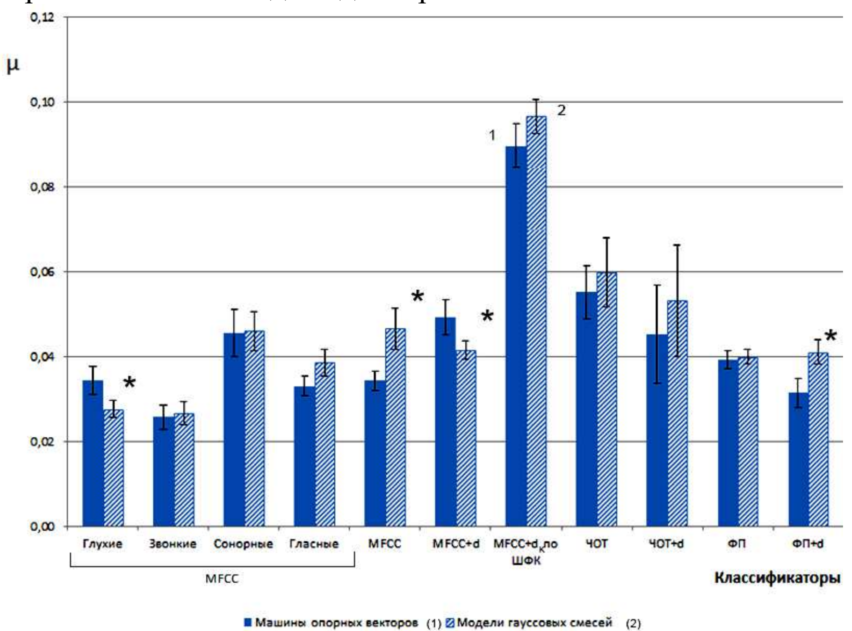


Рисунок 2 – Зависимость весовых коэффициентов (μ) от количества дикторов (n) «слабых» классификаторов с использованием в качестве акустических признаков частот основного тона (1, 1') и формантных пиков (2, 2'), применяя гауссовы смеси (1, 2) и машины опорных векторов (1', 2') в качестве методов классификации

Относительная оценка позволяет увидеть вклад отдельных «слабых» классификаторов в сравнении друг с другом (рис. 3). В качестве показателя тут применен усредненный весовой коэффициент бустинга после обучения на выборках от 2 до 50 дикторов, по 2 отрезкам произношений каждого диктора.



где MFCC – мел-частотные кепстральные коэффициенты;
ЧОТ – частота основного тона; ФП – формантные пики;
k – кол-во классификаторов; d – вектор признаков, дополненный производными

Рисунок 3 – Сравнение весовых коэффициентов (μ) «слабых» классификаторов, полученных различными методами, в решающем правиле, построенном алгоритмом бустинга

Исходя из графика можно сделать следующие выводы:

– наибольший вклад в решающее правило вносят классификаторы, основанные на совокупности отдельных ШФК;

– вклад методов значительно не отличается при использовании одинаковых характеристик, исключения составляют лишь характеристики, помеченные *, где отличия между классификаторами на основе различных методов достоверны (р < 0,05).

Сравнение эффективности «слабых» классификаторов с решающим правилом, построенным алгоритмом бустинга на множестве тех же классификаторов показало превосходство бустинга по доле верно идентифицированных дикторов в среднем на 2,4 ± 1,1% (р < 0,05) при $10 \leq N \leq 50$.

Выводы

Предложена схема текстонезависимой идентификации диктора, включающая предварительную сегментацию по широким фонетическим классам (ШФК) и иерархическое структурирование базы моделей дикторов. Показано, что разбиение речи на участки, принадлежащие различным ШФК, и создание для каждого диктора множества моделей фонетических элементов, увеличивает качество идентификации на 2,5 ± 0,8% ($p < 0,05$) по сравнению с моделями без предварительной сегментации.

1. Иерархическое структурирование позволяет избежать полного перебора базы моделей дикторов, ускоряя процесс закрытой идентификации в 4-5 раз в зависимости от количества дикторов (при $20 \leq N \leq 50$). Выигрыш в скорости возрастает с увеличением количества моделей дикторов в базе.

2. Вклад рассмотренных методов классификации значительно не отличается при использовании одинаковых характеристик, исключения составляют лишь характеристики MFCC и формантные пики с производными.

3. Применение бустинга позволило улучшить вероятность верной идентификации.

В ходе численных исследований решающее правило на основе бустинга текстонезависимо идентифицировало до 100 ± 0,2% (р < 0,05) дикторов при их количестве N=10 и до 97,2 ± 0,9% (р < 0,05) при N=50, что превосходит результат, полученный с применением рассмотренных «слабых» классификаторов, в среднем на 2,4 ± 1,1% (р < 0,05) при $10 \leq N \leq 50$. С возрастанием количества моделей дикторов преимущество комбинации классификаторов над «слабыми» классификаторами становится значительным.

Развитием данной работы для построения робастной текстонезависимой системы идентификации диктора могут стать следующие направления исследований:

– исследование эффективности проведения процедуры верификации по созданным моделям;

– исследование характеристик канальных искажений с целью построения их представительной базы данных для компенсации динамических помех в речевом сигнале;

– исследование возможности оптимизации скорости идентификации с помощью оценки весовых коэффициентов, полученных при обучении алгоритмом бустинга.

Список литературы

1. Zhao H. Boosting Image Retrieving Result with GPS Sensor Information / Zhao, Hongwei and Wang, Zhen and Liu, Pingping and Gao, Shan and Nong, Zhimeng // Sensor Letters. – American Scientific Publishers, 2014. – V. 12, № 2. – P. 213-217.
2. Sascha E.A. Muenzing. DIRBoost – An algorithm for boosting deformable image registration: Application to lung CT intra-subject registration / Sascha E.A. Muenzing, Bram van Ginneken, Max A. Viergever, Josien P.W. Pluim // Medical Image Analysis. – Elsevier, 2014. – V. 18, № 3. – P. 449-459.

3. Первушин Е.А. Обзор основных методов распознавания дикторов / Е.А. Первушин // Математические структуры и моделирование. − 2011. − № 24. − С. 41-54.
4. Сорокин В.Н. Верификация диктора по спектрально-временным параметрам речевого сигнала / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – Т. 10, № 2. – С. 87-104.
5. Клименко Н.С. Разработка структуры текстонезависимой системы идентифи-кации диктора // Искусственный интеллект. – 2012. – №4. – С. 161–171.
6. Martinez, J. Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques / J. Martinez, H. Perez, E. Escamilla, M.M. Suzuki // Electrical Communications and Computers (CONIELECOMP) 22nd International Conference on 27-29 Feb. 2012. – IEEE, 2012. – P. 248-251.
7. Ali Ali Khalil A.A. Efficient speaker identification from speech transmitted over bluetooth based system / Ali Ali Khalil A.A., Mostafa Saad E.S., El-Nabi M.A., Abd El-Samie F.E. // Computer Engineering & Systems (ICCES), 2013 8th International Conference on 26-28 Nov. 2013. – IEEE, 2013. – P. 190-193.
8. Садыхов Р.Х. Модели гауссовых смесей для верификации диктора по произвольной речи / Р.Х. Садыхов, В.В. Ракуш // Доклады БГУИР. – 2003. – №4. – С. 95-103.
9. Bartlett P. Generalization performance of support vector machines and other pattern classifiers / P. Bartlett, J. Shawe-Taylor // Advances in Kernel Methods. MIT Press, 1998. 13 p.
10. Zhu, J. Multi-class adaboost / Zhu, J., Zou, H., Rosset, S., Hastie, T. // Statistics and Its Interface. – 2009. – Vol. 2. – P. 349-360.
11. Электронный ресурс: http://voxforge.org/ru/downloads.

References

1. Zhao H. Boosting Image Retrieving Result with GPS Sensor Information / Zhao, Hongwei and Wang, Zhen and Liu, Pingping and Gao, Shan and Nong, Zhimeng // Sensor Letters. – American Scientific Publishers, 2014. – V. 12, № 2. – P. 213-217.
2. Sascha E.A. Muenzing. DIRBoost – An algorithm for boosting deformable image registration: Application to lung CT intra-subject registration / Sascha E.A. Muenzing, Bram van Ginneken, Max A. Viergever, Josien P.W. Pluim // Medical Image Analysis. – Elsevier, 2014. – V. 18, № 3. – P. 449-459.
3. Pervusin E.A. Obzor osnovnyh metodov raspoznavaniya diktorov / E.A. Pervusin // Matematicheskie struktury I modelirovanie. − 2011. − № 24. − P. 41-54.
4. Sorokin V.N. Verifikatsiya diktora po spektralno-vremennym parametram rechevogo signala / V.N. Sorokin, A.I. Tsiplihin // Informatsionnye processy. – V. 10, № 2. – P. 87-104.
5. Klymenko M.S. Razrabotka struktury tekstonezavisimoy sistemy identifikatsii diktora // Iskustveniy intelekt. – 2012. – №4. – С. 161–171.
6. Martinez, J. Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques / J. Martinez, H. Perez, E. Escamilla, M.M. Suzuki // Electrical Communications and Computers (CONIELECOMP) 22nd International Conference on 27-29 Feb. 2012. – IEEE, 2012. – P. 248-251.
7. Ali Ali Khalil A.A. Efficient speaker identification from speech transmitted over bluetooth based system / Ali Ali Khalil A.A., Mostafa Saad E.S., El-Nabi M.A., Abd El-Samie F.E. // Computer Engineering & Systems (ICCES), 2013 8th International Conference on 26-28 Nov. 2013. – IEEE, 2013. – P. 190-193.
8. Sadyhov R.H. Modeli gaussovyh smesey dlya verifikatsii diktora po proizvolnoy rechi / R.H. Sadyhov, V.V. Rakush // Doklady BGUIR. – 2003. – №4. – P. 95-103.
9. Bartlett P. Generalization performance of support vector machines and other pattern classifiers / P. Bartlett, J. Shawe-Taylor // Advances in Kernel Methods. MIT Press, 1998. 13 p.
10. Zhu, J. Multi-class adaboost / Zhu, J., Zou, H., Rosset, S., Hastie, T. // Statistics and Its Interface. – 2009. – Vol. 2. – P. 349-360.
11. Electronic resource: http://voxforge.org/ru/downloads.

RESUME

M.S. Klymenko, I.G. Gerasimov

Research of Boosting Efficiency for Text-Independent Speaker Identification System

In the article, the characteristics and problems faced by developers of text-independent speaker identification system are analyzed. Also the typical scheme of this task is briefly considered.

Based on the analysis, proposed structure of text independent speaker identification system with the addition of a hierarchical database of speaker's models and automatic pre-segmentation of the speech signal into sections containing phonemes from different wide phonetic classes (WFK).

Suggested hierarchical model database will reduce the number of comparisons between the model and sample. Thereby it will significantly speed up the identification process. Using classification of speaker-independent segments will neutralize the difference between training and recognizable context and allow to create a set of speaker models obtained by different WPCs. This can significantly improve the efficiency of identification.

Gaussian mixture models are used for model classification and support vector machines are used for group background models in the hierarchy. Formant peaks, frequency of the pitch and mel-frequency cepstral coefficients are used as an acoustic features of phonemes. Each combination of method-feature formed in "weak" classifier. The set of obtained classifiers were trained by boosting algorithm. It results in a single combined decision rule.

A numerical study of the efficacy of boosting which is applied to the set of "weak" classifiers acoustic performed. Combined decision rule by an average of 2.4% higher than the result obtained using "weak" classifiers.

Статья поступила в редакцию 05.06.2014.